Experiences of Using the GGF SAML AuthZ Interface

Dr Richard Sinnott
National eScience Centre
University of Glasgow
ros@dcs.gla.ac.uk

Professor David Chadwick
Systems Security IS Institute
University of Salford
D.W.Chadwick@salford.ac.uk

Abstract

The BRIDGES project has been funded by the UK Department of Trade and Industry to develop a Grid infrastructure suitable for the research activities involved in the Wellcome Trust funded Cardiovascular Functional Genomics (CFG) project. The CFG project is investigating possible genetic causes of hypertension. Key requirements on this infrastructure are to link various distributed biomedical data sources together; to transparently address the different security requirements associated with those data resources, and develop tools for analysing and exploring those data sets. In this paper we discuss the security solutions that the BRIDGES team is pursuing through the first practical exploration of Global Grid Forum Security Assertion Markup Language (SAML) AuthZ interface to an authorisation infrastructure (PERMIS) using Globus Toolkit version 3 technology.

1. BRIDGES Background

Hypertension affects a quarter of the adult population in western societies and is the major cause of cardiovascular mortalities. It is believed that hypertension is caused by a combination of factors including both genetic and environmental influences. The CFG project [1] is investigating the causes of hypertension and involves five UK and one Dutch site. It is pursuing a strategy combining studies on rodent models of disease with studies of patients and population DNA collections. The project is a prime example of the large-scale computational problems associated with modern biology, with requirements to combine vast arrays of heterogeneous information about three species, human, mouse and rat. Currently however, many of the activities that the CFG scientists undertake in performing their research are done in a time consuming and largely non-automated manner often requiring navigation to many different data resources, web sites and following multiple links to potentially relevant information. Similarly, in their pursuit of novel genes and understanding their associated function, the scientists often require access to large scale compute facilities to analyse their data sets, e.g. in performing large scale sequence comparisons or cross-correlations between large biological data sources.

The Biomedical Research Informatics Delivered by Grid Enabled Services (BRIDGES) project [2] has been funded by the UK Department of Trade and Industry to directly address the needs of the CFG scientists and provide a thorough investigation of relevant technologies for this purpose. Specifically, BRIDGES will investigate the application of Open Grid Services Architecture – Data Access and Integration (OGSA-DAI) [3] and IBM's Information Integrator product [4] to deal with federation of distributed biomedical data.

A key requirement of the scientist and hence focus of the BRIDGES work is security. Broadly speaking, the CFG scientific data can be classified dependent upon its security characteristics into three groups: public data (with no/minimal security, e.g. publicly curated genomic databases); shared data (belonging to the CFG scientists/consortia, e.g. shared research data sets); private data (belonging to given CFG sites and unavailable to anyone else, e.g. personal medical records).

2. Security Considerations

The Grid infrastructure to be deployed by BRIDGES should address all of the security concerns and interlinking of the different data sets in as transparent, and user friendly a manner as possible. It is widely recognised that the existing security solutions adopted by the Grid community have scope for improvement [5]. Currently for example, the UK e-Science community in establishing the Level-2 Grid has focused largely on PKI [6] based *authentication* mechanisms for security using X.509 [7] certificates issued by RAL. Authentication as the establishment and propagation of a user's identity in a given system is a useful starting point for security, but should ideally be augmented with, at a minimum, *authorisation* capabilities concerned with controlling access to services based upon specific policies. There are numerous technologies which claim to provide levels of authorisation in the context of the Grid including CAS [8], Akenti [9], VOMS [10], VOM [11] and PERMIS [12].

Given that BRIDGES has focused upon Globus toolkit version 3 (GT3) [13] as the basis for implementation of the initial family of Grid services accessing and using the different data sets, it is important that authorisation capabilities were made available through this technology. The Global Grid Forum Security Authorisation working group have developed specifications of generic Security Assertion Markup Language (SAML) [13] APIs through which authorisation infrastructures can be accessed and used. This API specification was prototyped in the PERMIS and Globus Toolkit (version 3.3) toolkits – thus permitting SAML authorisation callouts from Grid services to authorisation infrastructures to be realised. This work thus represents the first exploration of the SAML AuthZ interface to generic authorisation infrastructures.

2.1 PERMIS

The PERMIS software realises a Role Based Access Control (RBAC) authorisation infrastructure. It offers a standards-based Java API that allows developers of resource gateways (gatekeepers) to enquire if a particular access to a resource should be allowed. PERMIS RBAC uses XML based policies defining rules, specifying which access control decisions are to be made for given VO resources. These rules include:

- definitions of subjects that can be assigned roles
- definitions of Sources of Authority (SOAs) - trusted to assign roles to subjects
- definitions of roles and their hierarchical relationships
- definitions of what roles can be assigned to which subjects
- definitions of target resources, and the actions that can be applied to them
- definitions of which roles are allowed to perform which actions on which targets
- the conditions under which access can be granted to roles.

Roles are assigned to subjects by issuing them with X.509 Attribute Certificate(s). A graphical tool called the Privilege Allocator (PA) has been developed to support this process. Once policies are developed they are signed and stored in an LDAP repository.

2.2 SAML AuthZ Specification

The SAML specification defines a number of elements for making assertions and queries regarding authentication, authorization decisions and attributes The SAML AuthZ specification defines a message exchange between a policy enforcement point (PEP) and a policy decision point (PDP) consisting of an *AuthorizationDecisionQuery* flowing from the PEP to the PDP, with an assertion returned containing some number of *AuthorizationDecisionStatements*

The *AuthorizationDecisionQuery* itself consists of

- *A Subject* element containing a *NameIdentifier* specifying the initiator identity
- A *Resource* element specifying the resource to which the request to be authorized is being made.
- One or more *Action* elements specifying the actions being requested on the resources

The GGF SAML profile specifies a *SimpleAuthorizationDecisionStatement* (essentially a granted/denied Boolean) and an *ExtendedAuthorizationDecisionQuery* that allows the PEP to specify whether the simple or full authorization decision is to be returned.

3. System Design and Initial Experiences

Figure 1 provides an overview of the system used to explore the SAML AuthZ interface in Bridges. The GT3-PERMIS extensions realising the GGF SAML AuthZ profile allows for authorisation at portal access and subsequent Grid service invocations to be supported. The portal is personalised to CFG scientists based on the policies that have been defined for them, i.e. their role, targets etc. These policies are accessed when users log-in. Thus scientists are restricted to seeing and using services that are appropriate based on their roles.

A typical scenario that the infrastructure supports is:

- The user requests access to the CFG portal;
- The access request results in a SAML query being raised to ensure that this user is authorised to access the portal (by ensuring an appropriate policy is available in the secure LDAP repository);
- If successful (the user is authorised), the portal is configured/personalised to display the services that are associated with that user;
- At this point, the user can invoke various services (they are entitled to use) – one of these in a syntenic relation visualisation service (SyntenyVista).
- Upon launching SyntenyVista (using WebStart technologies) the users can use data available in the repository (which itself provides an OGSA-DAI front end and exploits IBM Information Integator to integrate and where possible federate various remote public data resources);
- The user may then visually explore genomic data sets and potentially export these onto the high throughput computing resources ScotGrid for sequence similarity checking (BLAST) against other query sequences.

In the current implementation the usage of SyntenyVista offers direct visualisation of data sets available via the repository (from ensembl [15]). It is planned however that the user is restricted to seeing and visualising the data sets that they are entitled to see based upon their role within the CFG virtual organisation (VO), this applies also to the usage/invocation of GT3 based Blast services, i.e. that they will be restricted to those users ***and*** those data sets that meet appropriate security restrictions.

Work is on-going to address this issue with initial thoughts that this will be realised through mapping of the user role within the CFG VO (as extracted from the secure policy database) against specifically established user views of data sets available via the DB2 data repository. However one issue that has been encountered with the SAML AuthZ profile is the lack of granularity in how users might invoke actions. For example, different actions may or may not be allowed depending upon the data that they wish to access and potentially change. The SAML AuthZ profile does not currently allow actions to be distinguished based upon the parameters that might be associated with them. As a result, the GT3 based BLAST service cannot be restricted to BLAST those data sets that are appropriate to the invoker. Instead, the SAML AuthZ specification supports either a SecureGrid BLAST service or a non-secure BLAST service. Thus when the portal is personalised per user/role, it is not possible to distinguish the usage of individual operations,

e.g. to allow arbitrary invocations of actions where the data sets themselves might change.

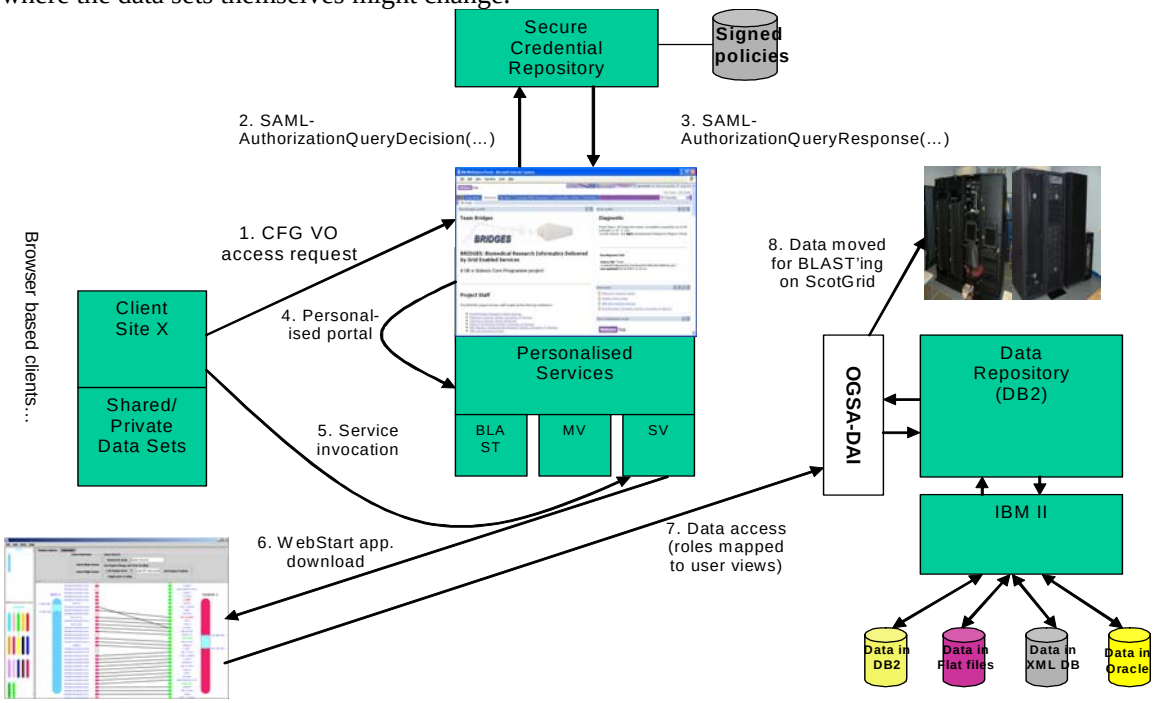

Figure 1: System Design and Usage Scenario

4. References

[1] Cardiovascular Functional Genomics project, http://www.brc.dcs.gla.ac.uk/projects/cfg/

[2] BioMedical Research Informatics Delivered by Grid Enabled Services (BRIDGES), www.brc.dcs.gla.ac.uk/projects/bridges

[3] Open Grid Service Architecture – Data Access and Integration project (OGSA-DAI), www.ogsadai.org.uk

[4] IBM Information Integrator, www.ibm.com

[5] E-Science Security Roadmap: Technical Recommendations v0.5, UK e-Science Security Task Form, draft executive summary v0.51

[6] Adams, C., Lloyd, S. (1999). "Understanding Public-Key Infrastructure: Concepts, Standards, and Deployment Considerations", Macmillan Technical Publishing, 1999

[7] ITU-T Rec. X.509 (2000) | ISO/IEC 9594-8, The Directory: Authentication Framework

[8] L Pearlman, et al., A Community Authorisation Service for Group Collaboration, in Proceedings of the IEEE 3rd International Workshop on Policies for Distributed Systems and Networks. 2002.

[9] M Thompson, et al., Certificate-Based Access Control for Widely Distributed Resources, in Proc 8th Usenix Security Symposium. 1999: Washington, D.C.

[10] VOMS Architecture, European Datagrid Authorization Working group, 5 September 2002.

[11] Steven Newhouse, Virtual Organisation Management, The London E-Science centre, **http://www.lesc.ic.ac.uk/projects/oscar-g.html**

[12] Privilege and Role Management Infrastructure Standards Validation project www.permis.org

[13] Globus toolkit, www.globus.org/toolkit

[14] P Hallem-Baker, E Maler, Assertions and Protocol for the OASIS Security Assertion Markup Language (SAML), v1.0 Specification. 31 May 2002. http://www.oasis-open.org/committees/security/#documents

[15] EMBL-EBI European Bioinformatics Institute, http://www.ebi.ac.uk/ensembl/,